CERTainty: Detecting DNS Manipulation at Scale using TLS Certificates

Elisa Tsai[*] Deepak Kumar[†] Ram Sundara Raman[*] Gavin Li[*] Yael Eiger[*] Roya Ensafi[*]
[*]Censored Planet, University of Michigan [†]Stanford University

ABSTRACT

DNS manipulation is an increasingly common technique used by censors and other network adversaries to prevent users from accessing restricted Internet resources and hijack their connections. Prior work in detecting DNS manipulation relies largely on comparing DNS resolutions with trusted control results to identify inconsistencies. However, the emergence of CDNs and other cloud providers practicing content localization and load balancing leads to these heuristics being inaccurate, paving the need for more verifiable signals of DNS manipulation. In this paper, we develop a new technique, CERTainty, that utilizes the widely established TLS certificate ecosystem to accurately detect DNS manipulation, and obtain more information about the adversaries performing such manipulation. We find that untrusted certificates, mismatching hostnames, and blockpages are powerful proxies for detecting DNS manipulation. Our results show that previous work using consistency-based heuristics is inaccurate, allowing for 72.45% false positives in the cases detected as DNS manipulation. Further, we identify 17 commercial DNS filtering products in 52 countries, including products such as SafeDNS, SkyDNS, and Fortinet, and identify the presence of 55 ASes in 26 countries that perform ISP-level DNS manipulation. We also identify 226 new blockpage clusters that are not covered by previous research. We are integrating techniques used by CERTainty into active measurement platforms to continuously and accurately monitor DNS manipulation.

1 INTRODUCTION

Due to a lack of encryption, DNS traffic is easy to manipulate, reroute, and hijack. DNS manipulation is a common technique used by censors and other adversaries to prevent users from reaching restricted Internet resources [14, 23, 38]. Conceptually, identifying DNS manipulation is straightforward and entails verifying the legitimacy of resolved IP addresses. However, in reality, detecting DNS manipulation on the global stage is more challenging due to website localization effects, differences in censor behaviors, and a dearth of clear signals of manipulation.

To address these challenges, prior work has proposed a myriad of detection mechanisms, most of which rely on comparing DNS resolutions and corresponding metadata collected through trusted *control* DNS resolvers with those collected through *test* DNS resolvers that are suspected of performing DNS manipulation [20, 24, 26, 45, 46, 66, 68, 76]. Such heuristics have been deployed by longitudinal censorship measurement platforms including OONI [20] and Censored Planet [66], providing open-access data to thousands of researchers in the Internet freedom community to identify and report censorship events. Unfortunately, the rise in popularity of CDNs and cloud providers, anycast-based routing, load-balancing, DNS misconfiguration, and localization has led DNS resolutions to often be unpredictable, significantly hampering the accuracy and usefulness of the "test metadata vs. control metadata" strategy proposed in previous work. These issues have caused measurement platforms to, in some cases, wrongly flag instances of DNS manipulation or completely overlook instances of manipulation [19, 55, 76]. Indeed, we show in this paper that more than 72.45% of the DNS manipulation detected using state-of-the-art heuristics are false positives. Due to the far-reaching implications of censorship measurement, it is crucial that the identification of DNS manipulation is performed accurately.

In this paper, we propose a novel technique, *CERTainty*, to detect DNS manipulation by utilizing a widely adopted trust infrastructure: *TLS certificates*. *CERTainty* relies on the fact that valid TLS certificates for a domain can only be issued by its owner, and DNS manipulation is performed by an in-network adversary such as an ISP that does not own the domain. *CERTainty* fetches TLS certificates from the IP addresses returned during the DNS resolution and examines the validity of these certificates for the requested domain. To do so, we equip *CERTainty* to validate certificates with a well-known root store and consider cases where TLS certificates are mismatched or untrusted. We evaluate our technique by matching responses with HTML blockpages as ground truth and identify that almost all cases of certificate invalidity can be mapped back to true instances of DNS manipulation. Moreover, we use information from certificates and HTML blockpages to attribute DNS manipulation and find who implements it.

We evaluate our research with previous studies that have used TLS certificates and state-of-the-art heuristics that rely on control metadata to detect DNS manipulation [53, 62, 66], and find that techniques used previously are highly flawed. We discover that previous work does not properly consider the effects of hostname matching, certificate misissuance, and captive portals, leading to poor performance of TLS certificate-based detection. Moreover, we observe that the dynamic behaviors of CDN and website content localization cause 72.45% of DNS manipulation cases detected using "test vs control" heuristics to be false positives. Moreover, these heuristics also fail to detect 9.70% of the true cases of DNS manipulation identified by *CERTainty*.

We show how certificate validation and blockpage fingerprint matching provide a holistic view of DNS manipulation, allowing not only detection but attribution. Globally, *CERTainty* identifies 17 TLS proxy vendors in 52 countries, including products such as SafeDNS, SkyDNS, and Fortinet. We discover 7 commercial

Proceedings on Privacy Enhancing Technologies YYYY(X), 1–16

https://doi.org/XXXXXXX.XXXXXXX

	Measurement Range	Consistency								Verifiable Signals		
		IP	HTTP	Cert	ASN	ASNa	PTR	TTL	Thres	Cert	Page	Manual
OONI (2012) [20]	Global (>200 countries)	●			●		●				●◇	
Censored Planet (2020)[66]	Global (221 countries)	●	●	●	●	●	●					
IClab (2020)[24, 45] ⋆	Global (62 countries)	●			●				●		●◇	
Yadav et al. (2018) [76]	India	●			●							●
Iris (2017)[53] ⋆	Global (151 countries)	●	●	●	●	●	●			●		
REMeDy (2017)[68]	Local ISPs	●			●			●				
UBICA (2015)[2, 3]	Pakistan, South Korea and Italy	●									●	
Verkamp et al. (2012) [72]	Global (11 countries)	●					●					

Table 1: DNS manipulation detection heuristics—We omit platforms that deploy country-specific measurement techniques based on the targeted DNS manipulation systems e.g. China [26, 37, 46] and Iran [5]. The ⋆ symbol indicates the platforms that consider RCODE and private IPs. All the heuristics mentioned are for DNS manipulation with public IPs returned. The consistency heuristics contain control-matching IP, HTTP hash, certificate hash, AS number, AS name, PTR record, packet TTL field, and tuned threshold (for the number of domains mapping to the same IP). The verifiable signals heuristics include the certificates (with and without SNI) fetched from the resolved IPs, web page (length and blockpages ◇), and manual analysis.

products that are deployed in more than one country and find that these products return a small pool of common tampered IP addresses across different countries, providing a good avenue for both detection as well as circumvention. *CERTainty* also detects 55 ASes in 26 countries with ISP-level DNS manipulation, which includes both countries with well-known DNS blocking systems (e.g. Russia [60] and China [25, 26, 39, 46, 74]), as well as countries that have not been studied in previous research (e.g. Nepal, Latvia, Poland, and Singapore). In both the commercial filtering product deployments as well as ISP-level deployments, we find a wide variety in the types of certificates and details in blockpages returned. To our best knowledge, this is the first report of DNS manipulation deployment (products and ISPs) on a global scale.

Our results highlight the advantages of using established trust mechanisms such as certificate validation and blockpage matching, not only in accurately detecting DNS manipulation, but also in gaining more knowledge about the adversaries performing such manipulation. We have integrated *CERTainty* into Censored Planet [66], a remote censorship measurement platform, and data generated by *CERTainty* is actively being published and utilized by hundreds of researchers. We are also actively working on integrating our techniques into other measurement platforms such as OONI [20]. We hope that our techniques bring improved accuracy and rigor to the continued monitoring of DNS manipulation attempts.

2 BACKGROUND

In this paper, we define DNS manipulation as the phenomenon where a network adversary — such as a censoring authority — manipulates DNS responses to prevent a user from accessing legitimate content for the name requested in the DNS query. DNS manipulation has been studied both in country-specific contexts [5, 7, 9, 17, 26, 42, 46, 60, 63, 76] and with global measurements [20, 45, 62, 66, 68], revealing its diverse and decentralized nature. Countries like Pakistan use a nonzero Response Code (RCODE), e.g. NXDOMAIN, to deny access to blocked domains [42]. Others, like Russia, manipulate DNS responses at the ISP level and redirect users to blockpages [60]. A few countries perform manipulation in a more centralized manner, deploying their national firewall at the Internet backbone, and either return private IPs [5, 75] or a pool of designated IPs [26, 46]. The technological barrier to deploying DNS manipulation on a national scale has fallen, as middlebox vendors from nations with developed filtering technologies increasingly export their wares to those without them, making censorship implementation simple [13, 59, 60, 73].

***Measuring DNS Manipulation*:** As established in previous work [20, 45, 53, 66], the DNS manipulation we aim to detect contains two scenarios: (1) the resolution is unsuccessful either because an in-path adversary drops the connection, or because poisoned or tampered DNS resolvers return nonzero RCODE for domains on the blocklist, and (2) resolved IPs do not host the requested domains e.g. private IPs or IPs hosting a blockpage stating that access is blocked.

***DNS Manipulation Detection Heuristics*:** As shown in Table 1, most censorship measurement platforms such as OONI [20], Censored Planet [66], IClab [45], Iris [53], REMeDy [68] and UBICA [3] incorporate a "test vs. control" strategy, with requests to trusted resolvers acting as control *ground truth*. However, in the current Internet landscape that contains localization effects, it is challenging to ensure that such controls identify all intended resolutions. Therefore, when comparing IP addresses to control measurements is inconclusive, measurement platforms use a variety of other control-matching heuristics to determine whether DNS resolution is correct. These heuristics often fall into two categories: (1) consistency-based heuristics and (2) verifiable signals.

***Detecting DNS Manipulation through consistency-based heuristics*:** The design philosophy of consistency-based heuristics is to confidently identify *unmanipulated* DNS responses. If the IP address or *any* of the other metadata matches with the corresponding metadata in the control group, the DNS response is tagged as unmanipulated. The heuristics were designed based on the insight that, behind the same domain name, there are typically shared infrastructural signals even if the exact IP address is different. For instance, Pearce et al. showed in 2017 that heuristics such as the AS number and name, HTTP content hash, HTTPS certificate hash, and

PTR records serve as good consistency heuristics [53]. Censored Planet, an active censorship measurement platform, also uses similar heuristics [66].

Despite their usefulness, we show in this paper that consistency-based heuristics such as the ones introduced in Pearce et al. [53] face a number of challenges that make them unsuitable for large-scale DNS manipulation detection (see more discussion in §5.2). First, consistency-based heuristics rely on the availability of infrastructure metadata, such as AS information, which is not always accurate [56]. Second, consistency-based heuristics result in a number of false negatives due to the fact that legitimate content and adversaries could both use the same infrastructure, such as hosting information on a CDN. Finally, consistency-based heuristics also result in a large number of false positives, since legitimate content could be hosted in different infrastructures in different regions. Because of these challenges, in this paper, we instead rely on designing verifiable signals of DNS manipulation.

***DNS Manipulation Detection through verifiable signals*:** An alternate approach to detecting DNS manipulation is to use independent signals that can indicate whether the IP address returned during DNS resolution provides legitimate content. For instance, if injected or poisoned IPs redirect traffic to a blockpage citing the reason for blocking, we view it as a very strong signal of DNS manipulation. Previous work has used a range of clustering techniques to identify blockpages. Dalek et al. created signatures for 4 URL filter vendors in 2014 [12]. In 2014, Jones et al. utilized page length and term frequency vectors [31] to discover blockpages. In 2020, Niaki et al. used textual similarity and HTML structure similarity to cluster potential blockpages [45]. In the same year, Sundara Raman et al. created blockpage fingerprints on the application layer for more than 90 vendors and actors [59]. All of these previous studies manually curated the fingerprints of the blockpages. Human identification remains the primary mechanism to identify the unique parts of blockpages of various domains.

Pearce et al. [53] used certificates (both with and without SNI) to identify unmanipulated DNS resolution. While they reported below-average performance in the use of certificates for detecting DNS manipulation, we find that their approach is error-prone (§5). We instead demonstrate in §4 that with proper consideration of hostname-matching, certificate misissuance, and captive portals removal, the validity of certificates with domain SNI can serve as an effective detector of DNS manipulation. Using certificate validation, we not only pinpoint the signals and the corresponding actors of DNS manipulation, but also detect more covert forms of DNS manipulation where no clear signals of blocking are shown to the user, often granting the adversaries plausible deniability [29].

3 DATA

We leverage open-access global DNS measurement data provided by Censored Planet [66]. Censored Planet performs measurements to thousands of open DNS servers longitudinally, and uses consistency-based heuristics such as AS number and name, HTTP content hash, and HTTPS certificate hash to determine DNS manipulation. We propose two novel techniques for improving Censored Planet's DNS manipulation detection, both of which involve making an HTTP(S) connection to the IP addresses returned during the DNS process: 1) When a TLS Client Hello message with the appropriate Server Name is sent to resolved IP addresses, *CERTainty* checks the validity and correctness of the returned certificates, and 2) *CERTainty* clusters and identifies blockpages, and determines whether the web page returned during the HTTP request matches our list of expert-curated blockpage fingerprints.

Fetch Page Status	Control Pages	Test Pages
Has TLS cert and HTTP page	5,898 (94.18%)	530,170 (83.71%)
HTTP Page Only	234 (3.74%)	50,287 (7.94%)
Connection error for both HTTP and HTTPS	130 (2.08%)	52,884 (8.35%)
Total	6,262	633,341

Table 2: Page Fetch Result Distribution—for control group and test group respectively.

3.1 Censored Planet DNS Data

In this section, we describe the global DNS resolution data collected by Censored Planet[66], an open-access remote website accessibility measurement platform. Note that the techniques we propose in this paper can be integrated into both in situ (inside a country of interest, running on VPNs or volunteer machines) as well as remote (utilizing open DNS resolvers in the country of interest) measurements. Since remote measurements offer the advantage of increased scale and coverage, we deploy and evaluate our detection techniques on remote measurement data in this paper.

***Trusted Control Resolvers*:** Censored Planet re-implements the consistency-based heuristics from Pearce et al. [53], and it compares DNS resolution results collected from a set of *trusted control resolvers* to the results of the test DNS resolvers. Censored Planet utilizes trusted, load-balanced, public open resolvers such as those operated by Google, Cloudflare, and UltraDNS as controls.

***Test Resolvers*:** Censored Planet leverages Censys [15] to locate open resolvers [15, 16], and filters these resolvers specifically to target *only* ones that can be identified as infrastructure DNS resolvers by retaining resolvers with PTR records containing `ns[0-9]+`, `nameserver[0-9]*`, `.telecom` or `.isp.`. This criterion is important, as erroneously leveraging user-controlled resolvers in censored countries can lead to unwanted legal or government action against citizens [21, 36, 77]. This process yields more than 25,000 open resolvers, spanning more than 200 countries.

***Domain Test List*:** The domains tested by Censored Planet are a combination of (1) the Citizen Lab Global Test List [35], which is a curated list of URLs intended to enable global censorship measurements. As of Nov 2022, the list has 1,598 domains and (2) 500 top domains from the Tranco 1M list [18], which is a list of popular domains updated daily.

***Censored Planet Data Characteristics*:** In this paper, we utilize Censored Planet [66] DNS data collected twice a week from May 16, 2022, to Nov 30, 2022. Each measurement snapshot consists of sending queries to an average of 25,943 open resolvers for 2,000 domains selected as mentioned above, resulting in over 2.93 billion lines of DNS resolutions in total. Among the 2.93 billion DNS

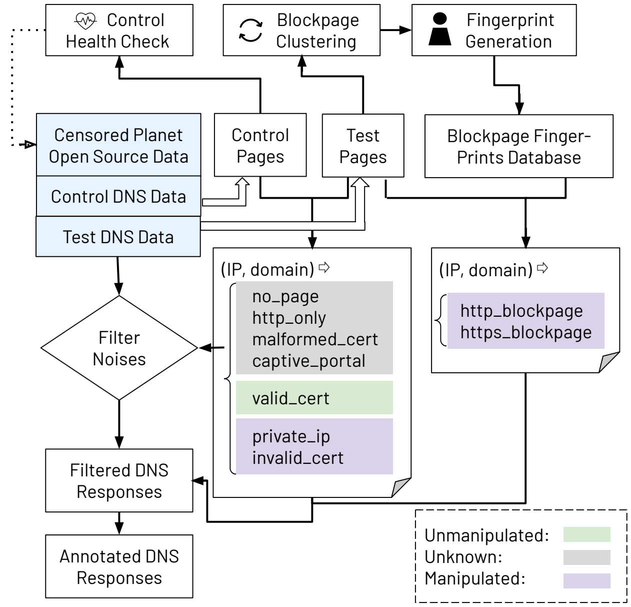


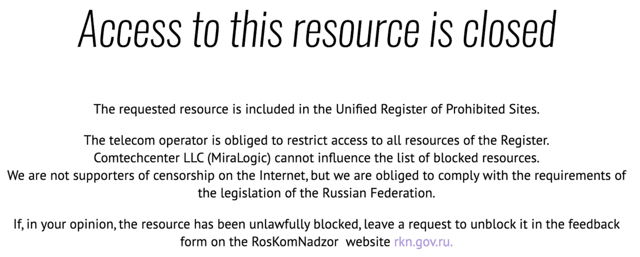
Figure 1: The DNS manipulation detection, noise removal, and annotation pipeline of *CERTainty*—The blue part indicates data retrieved from Censored Planet. The rest are data collected for this study.

resolutions, 96.87% succeed in getting a DNS response (i.e. do not experience a timeout), 0.006% receive a nonzero RCODE, and 93.45% of queries have at least one IP returned. For connection errors and nonzero RCODE, determining if the domain is accessible in a given region is a simple boolean check. However, when one or more IPs are returned, things are more complicated since many domains own myriad points of presence across the globe, hosted by various CDNs, which the control DNS resolutions fail to cover. This set of DNS resolutions is the main focus of this paper.

3.2 Fetching Content From Resolved IPs

As shown in Figure 1, in order to perform certificate validation and blockpage matching, *CERTainty* performs HTTP(S) requests for all the DNS resolution pairs, determining if the result would appear as censorship to a user. The domain is populated into the HTTP Host Header and the SNI extension for HTTP and HTTPS requests respectively. The unencrypted HTTP header and HTML pages are used for blockpage clustering to generate fingerprints. The certificate chains collected by the HTTPS requests are collected for certificate validation. We perform all page fetching measurements from a vantage point in North America.

***Measurement Characteristics*:** We issue 31.17 million HTTP(S) page requests for 2.93 billion DNS resolutions, over the course of 6 months. For each scan, there are on average 6,639,603 unique DNS resolution pairs. We perform HTTP (port 80) and HTTPS (port 443) page fetches to these DNS resolution pairs. As shown in Table 2, we successfully connect to port 443 and obtain a certificate and an HTTP response in over 83.71% of `(IP, domain)` pairs from our test cases, and obtain an HTTP page over port 80 in another 7.94% of cases. In about 8.35% of the cases, we see a TCP-level connection error for both HTTP (port 80) and HTTPS (port 443) requests.


Access to this resource is closed

The requested resource is included in the Unified Register of Prohibited Sites.

The telecom operator is obliged to restrict access to all resources of the Register.
Comtechcenter LLC (MiraLogic) cannot influence the list of blocked resources.
We are not supporters of censorship on the Internet, but we are obliged to comply with the requirements of the legislation of the Russian Federation.

If, in your opinion, the resource has been unlawfully blocked, leave a request to unblock it in the feedback form on the RosKomNadzor website rkn.gov.ru.

Figure 2: Example of blockpages *CERTainty* detects – A translated blockpage from MiraLogic, a Russian telecom company

Category	Product	National	ISP	Corporation	Unknown	General
Count	29	92	46	14	15	30

Table 3: Blockpage fingerprint distribution: The number of unique blockpage fingerprints under different categories.

The HTTP(S) connection errors *could* be a signal of DNS manipulation that requires further investigation. For example, if TCP resets are observed, it could be the adversaries resetting TCP connection for blocked domains (although this is unlikely since our measurement infrastructure is within a University network). More likely, these are cases where domains are not active on port 80 or port 443, or these domains geoblock requests from our measurement infrastructure. We discuss the connection errors in §7, and provide a case study in §6.4.

3.3 Blockpage Fingerprint Generation

In order to capture signals of overt censorship where a blockpage is served, *CERTainty* fetches HTTP response headers and HTML pages from the IPs returned by the control resolvers and test resolver. Previous work has investigated different methods to identify blockpages, utilizing clustering techniques based on the similarity of page length [31, 45], term frequency vectors [31], HTML structure [45] and the screenshot of the returned pages [59].

All of these blockpage clustering techniques are followed by a step to manually create blockpage fingerprints and appropriate labels for them. Constructing meaningful fingerprints manually is a widely accepted practice of blockpage detection. Tedious as it seems, these blockpage fingerprints provide valuable insights for the research community to track the scope, scale, and evolution of content-based censorship.

In this paper, we integrate publicly available blockpage fingerprints [54] generated previously from HTTP and HTTPS connection interference data [59], and complement it with 226 new blockpage fingerprints generated from Censored Planet DNS data. We observe that clustering the pages in the HTTP response based on page length and HTML structure works the best. Figure 2 is an example of ISP-level blockpage discovered by our semi-automatic blockpage detection. In total, 26 countries' ISP-level blockpage are discovered, including countries that are not covered by previous research e.g. Nepal, Latvia, Poland, and Singapore.

We manually verify each blockpage cluster in order to remove false positives. The presence of blockpages is relatively stable in our dataset. For over 8 months' time (from March 2022 to Nov 2022), only 21 new potential clusters are observed. Therefore future manual efforts for generating blockpage fingerprints are manageable. We craft the blockpage fingerprints into 6 categories as shown in Table 3, following the convention set by previous work [59]:

(1) *Commercial product*: commercial middleboxes e.g. OpenDNS [11], NextDNS [44], and OneDNS [50].
(2) *National-level DNS manipulation*: e.g. Indonesia blockpages that contains "Internet Positif (Positive Internet)".
(3) *ISP-specific blockpages*: blockpages which specify the ISPs who configure the blockpages e.g. Fig 2.
(4) *Corporational or institutional DNS manipulation*: e.g. blocking implemented by companies and universities.
(5) *Unknown*: blockpages that we do not have enough information to attribute the deployer. Each such fingerprint is annotated with the country of origin.
(6) *General*: e.g. "This Site Has Been Blocked" in the title.

Censored Planet has already incorporated HTTP(S) page fetch, certificate validation, and blockpage fingerprint matching into their weekly measurement process. Therefore, the blockpage fingerprints are also used to conduct health checks for control resolvers, allowing future control resolver list expansion. The annotated fingerprints are open-sourced for the community. We utilize blockpage fingerprints to serve as ground truth in evaluating certificate validation in §4.

3.4 Noise Removal

Censored Planet issues about 50 million DNS queries to more than 25,000 open resolvers across the globe. We add an extra filtering stage to exclude resolvers and IPs with erroneous behaviors. For example, if a resolver is returning erroneous responses for all of the queried domains, it is highly unlikely because of DNS manipulation. Instead, the cause is possibly misconfiguration or misguided NATs and firewalls [33]. Therefore, we exclude DNS responses from resolvers whose DNS responses either only contain timeouts, or have a nonzero RCODE, an empty list of IP addresses, private IP addresses, or the same set of IP addresses (possible captive portals) for *all* the queried domains.

Previous platforms proposed by Sundara Raman et al. [66] and Pearce et al. [53] ignore connection errors and nonzero RCODE responses entirely since many are not due to DNS manipulation. We deploy a more conservative filtering strategy to find signals of DNS manipulation in these responses. In §A.1, we show how proper filtering can help us capture signals of DNS manipulation within the NXDOMAIN RCODE.

3.5 Ethics

In this paper, we use DNS measurement data collected by Censored Planet, which follows best practices in selecting measurement vantage points and conducting measurements [66]. Censored Planet only selects DNS resolvers belonging to the Internet infrastructure such as nameservers for performing measurements. As highlighted by previous work, this is an attempt to eliminate any use of resolvers or forwarders owned by individuals [16, 52, 53, 66, 71]. This step significantly minimizes the risk, because the risk posed to administrators with more skills and resources to understand the traffic is lower than the risk posed to end users. We also set up WHOIS records and a web page served from port 80 of our measurement machine that indicates that our HTTP and HTTPS measurements are part of a research project and offer administrators the option to opt out of our scanning. We did not receive any inquiries or complaints over the period of 6 months.

4 USING CERTIFICATE VALIDITY TO MEASURE DNS MANIPULATION

Prior work in DNS manipulation detection has not incorporated TLS certificate validity into DNS manipulation detection properly [53]. However, the presence of a valid certificate (i.e., one trusted by a known root store and containing the correct hostname) is a strong signal that the application-layer connection to a server (HTTPS) is legitimate. In this section, we examine how we can use certificate validation as a proxy to evaluate the presence of DNS manipulation.

For the scope of this study, we consider a certificate to be valid if two criteria are met: (1) the certificate chains to a trusted root in the Mozilla NSS Root Store (used by Mozilla Firefox) by OpenSSL [51], and (2) the hostname in the certificate (either in the common name or the subject alternative name) matches the domain we are attempting to reach, following the rules as specified in RFC 6125 [27]. We note that *CERTainty* regards expired certificates as invalid, as we are using OpenSSL to verify the chain. Our approach is similar to the one followed by a browser attempting to validate the authenticity of a domain. At a high level, we consider any connection that returns a valid certificate to be unmanipulated and use other signals (e.g., blockpage fingerprints) to link certificate invalidity to DNS manipulation. We utilize 662 blockpage fingerprints, from both publicly available blockpages [59] and blockpages *CERTainty* detected.

We consider four distinct cases in certificate validation when identifying DNS manipulation where the certificate we obtain from control resolutions are valid, and one case where the certificates we obtain from control resolutions are invalid.

Our analysis is based on 12 snapshots over the course of 6 months. We do not identify significant differences among the snapshots—-we identify 8 new invalid certificate issuers and 13 new blockpage fingerprints in 6 months. Therefore, the analysis in this paper is based on one snapshot in Nov 2022. We discuss the longitudinal aspect of DNS manipulation in §7.

4.1 Case 1: Valid Certificates

We view the presence of a valid certificate for the requested domain as a strong signal that the IP address is *not* manipulated. Indeed, we note that none of the HTML pages returned with a valid certificate match a known blockpage fingerprint. We receive a handful (5.34%, 315 out of 5,898 total requests) of 403 error pages from CDNs due to cases such as geoblocking, where HTTP requests from our measurement infrastructure are blocked due to their location. However, the certificate for the requested domain is still valid in these cases, suggesting that certificate validation can help to eliminate the effect of geoblocking based on the vantage point chosen for measurements.

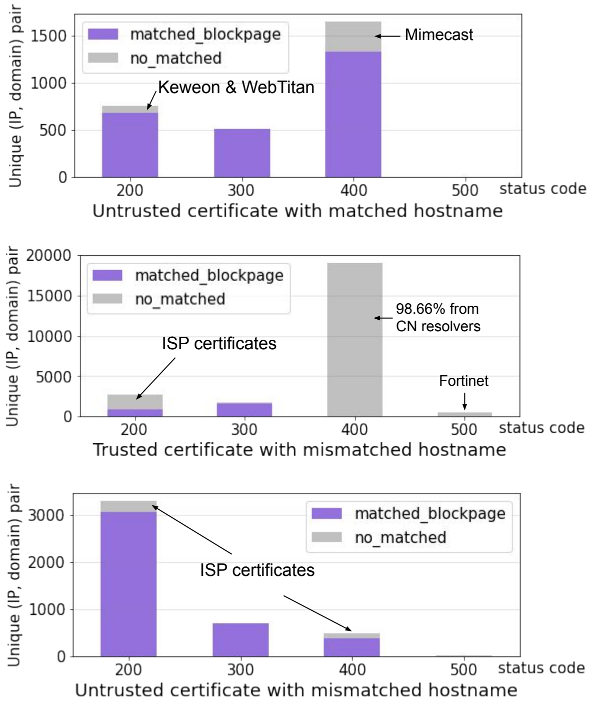


Figure 3: Blockpage fingerprint matching and control certificate matching for HTTPS responses with invalid certificate—The purple part denotes the certificates with an identified blockpage. The grey part denotes the proportion of corresponding HTML pages that do not match any known blockpage fingerprint. IPs of captive portals and invalid certificates that match with misissued control are removed as described in Sec. 3.4 and Sec. 4.5.

4.2 Case 2: Untrusted Certificate With Matched Hostname

If an untrusted certificate is returned with a matching hostname for a request, we mark the request as potential DNS manipulation. To confirm this categorization, we check our blockpage fingerprints against the pages returned during the HTTP request. As shown in Figure 3-(1), we observe that 86.25% (2,521 out of 2,923) of the untrusted certificates with a matching hostname come with an identified blockpage. For the other 13.75%, only 3 TLS product vendors are identified using the certificate issuer field. Keweon [32] and WebTitan [67] return an empty 200 OK HTTP page along with the certificates, making it impossible to craft a blockpage fingerprint for them. Mimecast [41] returns a general 404 error page. This shows that information extracted from certificates can be critically informative when no blockpages are presented. Certificates with untrusted root and matching hostname are strong signals of TLS proxies. Notably, two vendors—SkyDNS and SafeDNS—return pages with 451 status code (“Unavailable For Legal Reasons“), indicating that they are used by ISPs or governments. In total, we discover 12 such TLS proxy vendors and report our results in §6.1.

4.3 Case 3: Trusted Certificate With Mismatched Hostname

When a trusted certificate is returned with a mismatched hostname, we consider this to be a potential sign of DNS manipulation. Exploring these cases, we observe this behavior to be largely driven by ISP-level blocking. Of the requests made in this category, 10.48% (2,518 out of 24,029) of them match a blockpage fingerprint, as shown in Figure 3-(2). For requests that return 400+ status codes, 98.66% (18,825 out of 19,079) of these requests are returned by Chinese open resolvers, and those IPs typically belong to large entities like Facebook (66.30%, 12,481 out of 18,825), Twitter (29.10% 5,478 out of 18,825), Cloudflare (3.36% 632 out of 18,476), and other blocked CDN services e.g. Fastly and Akamai (less than 0.08%). Our observations align with prior China-focused studies [25, 26, 39, 46, 74] that suggest China's national Firewall (the GFW) returns IP addresses of large US-based companies to DNS queries of blocked content. The rest are mostly from Canadian Shield [10], a Canadian TLS middlebox vendor. For the 1.18% (543) requests that return a 500+ status code, we observe that 34.62% are returned by Chinese resolvers and point to the IP address of a large entity mentioned above. The remaining 62.98% certificates are issued by Fortinet, a well-known middlebox product vendor.

For requests that return a 200 status code, 34.28% (917 out of 2,675) of the returned webpages match a known blockpage fingerprint. We manually investigate the 65.72% of 200 status code webpages that we did not identify as blockpages, as well as the certificates that we collected for these requests. We identify 88.22% are ISP-issued certificates coming with “blockpage“, “allownet“ or “illegal“ in the certificate common name. For example, we see a certificate signed by ‘‘illegal.mdes.go.th‘‘ without meaningful page content, which is the Ministry of Digital Economy and Society of Thailand. This again proves that blockpage information alone is not enough for DNS manipulation detection. More ISPs that return informative certificates without blockpages can be found later in Table 10. The rest are instances of phishing, where we see that traffic is diverted to advertisement websites. Finally, for cases where a 300 status code is returned, we observe a large number of cases where domains have misconfigured TLS certificates and discuss these further in §4.5.

4.4 Case 4: Untrusted Certificate With Mismatched Hostname

An untrusted certificate with a wrong hostname is a very strong signal that the returned IPs do not host the requested domain, and is therefore a potential signal for DNS manipulation. We observe 92.31% (4,167 out of 4,514) of the pages match with a blockpage fingerprint. Upon further manual investigation for the 2.57% of unidentified pages, we find that this category of certificate likely originates from a misconfiguration. Certificates with the common name e.g. “testexp“, “test“ and “Plesk“ are returned with blank pages. Only 9 IPs in our whole dataset host such certificates. For the general 400+ error pages, we see certificates issued by ISPs in a few countries e.g. Singapore, Columbia, and Russia. The information in the certificate (e.g. common name ‘‘*.block.msm.ru‘‘) highlights these are cases of DNS manipulation even when there is no explicit blockpage.

Matched Heuristics	HTTP hash	Cert hash	ASN	AS name	CDN
Count	372	460	10,388	10,384	11,937
Percentage	3.12%	3.85%	87.02%	86.99%	100.00%

Table 4: False negatives introduced by consistency heuristics—The table shows the number and percentage of total cases where each consistency heuristic shows a match between test and control experiments, but our technique indicates DNS manipulation due to an invalid certificate or presence of blockpage.

Summary: Our results give us strong confidence that certificate validation is an effective proxy to detect DNS manipulation. It provides a venue to perform quick automated detection of DNS manipulation, reveals critical information when the middleboxes choose not to return blockpages, and can even help us discover covert DNS manipulation (more in §6.3).

4.5 Case 5: Invalid Control Certificate (Misissuance)

In order to use certificate validity as a proxy for detecting DNS manipulation, we need to account for certificates that would be invalid even in a control setting, as invalid certificates are common on the Internet [1, 4, 34, 70]. These "control certificates" serve as ground truth for the case where the certificates are invalid because of deployment errors from domain administrators.

We see 1.2% (72 out of 5,898) invalid certificates among all unique control certificates collected. This means the certificates for those domains are not issued correctly by default and the presence of invalid certificates for those domains is not necessarily a signal of DNS manipulation. We fetch TLS certificates regardless of the website's HTTP-based by default. We note that the number of domains with misconfigured TLS is much lower than what previous studies have found [53], possibly due to increased HTTPS adoption. We see a good proportion of .mil domains failing certificate validation such as `www.dtic.mil`. The US military (DoD) websites utilize Federal PKI, which is not trusted by most root stores [48]. Other misconfigured domains (e.g. `www.freeexpression.org` and `www.kcna.kp`) either have invalid certificates because of mismatching hostname, or untrusted root CA. We do not consider cases where domains have invalid control certificates as a signal of DNS manipulation.

4.6 Origin of Invalid Certificates

In this paper, we use the validity of certificates as a proxy to detect DNS manipulation. Theoretically, an adversary such as a censor that is in-path can inject invalid certificates by inspecting the HTTPS requests e.g. the SNI field in Client Hello. Thus, there is the possibility that we are misclassifying network manipulation at the TLS layer as DNS manipulation. However, note that this is unlikely since we are issuing HTTPS requests from a university network where no censorship is implemented instead of sending requests from inside a censored country where censorship can happen on different network stacks.

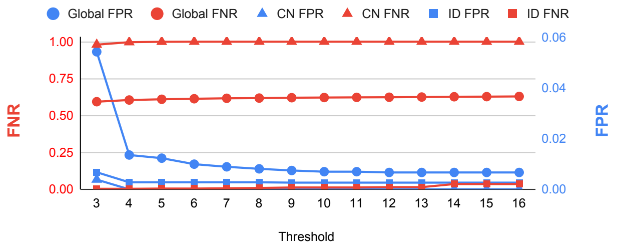


Figure 4: The Efficacy of the AS threshold heuristic—Generated by comparing statistical AS thresholding to the results of certificate validation and blockpage matching on the global scale, China and Indonesia, respectively.

Nevertheless, to confirm that the certificates we receive are indeed originating from the IP addresses received during DNS resolution, we perform TTL-limited traceroute tests for `(IP, domain)` pairs with invalid certificates, using methods developed in previous work [65]. We perform two TLS Hello requests for the control domain (`exmaple.com`) and the target domain, sending the requests with incrementing TTL values. Then we compare the control traceroute and test traceroute to determine where in the network the TLS response is originating from.

The results confirm our hypothesis. In all cases, we observe that the traceroute terminates in the same network (/24 subnet) as the endpoint IP address. Indeed, 93.24% (5,729/6,144) of traceroutes end within `+-1` of the hop where the control traceroutes end. Therefore, we are confident that the certificates are returned by the IPs obtained during DNS resolution.

5 EVALUATION

In this section, we assess the effectiveness of our method for identifying DNS manipulation using certificate validation and blockpage matching. As a baseline, we compare our technique directly against several heuristics proposed by the current state-of-the-art in two categories: (1) *verifiable signals* and (2) *consistency heuristics*, which compare potentially censored response data (e.g., IP, ASN, HTTP page hash) to trusted control responses.

5.1 Previous Verifiable Signal Heuristics

Some prior work leverage *verifiable signals* such as certificate validation [53], returned pages [20, 45, 76], and manual analysis [76]. We demonstrate how previous certificate validation is error-prone, and how our blockpage fingerprints enrich the body of knowledge for censorship detection.

***Certificate Validation*:** Pearce et al.'s [53] Iris technique established in 2017 checks whether the returned certificates for servers that support HTTPS are browser-trusted. For requests with SNI, the technique checks if they are for the correct IP addresses. However, in retrospect, this is no different from control IP matching. Moreover, the HTTPS ecosystem has changed a lot since 2017. In our data, only 0.12% of certificates have IP addresses in their common name or SAN, therefore this method is not applicable anymore.

For requests without SNI, Iris checks if the returned certificate contains the right domain name. However, many CDNs would return

No control hueristic matching				At least one control hueristic matching			
Comparison	*CERTainty* Result	Count	Percentage	Comparison	*CERTainty* Result	Count	Percentage
Same with *CERTainty*	Invalid Cert	95,624	13.98%	Contradict with *CERTainty*	Invalid Cert	11,097	0.13%
	HTTP Blockpage	15,492	2.27%		HTTP Blockpage	840	0.01%
Contradict with *CERTainty*	Valid Cert	495,532	72.45%	Same with *CERTainty*	Valid Cert	7,529,487	88.85%
Unconfirmed by *CERTainty*	HTTP Only	33,592	4.91%	Unconfirmed by *CERTainty*	HTTP Only	186,627	2.20%
	Connection Error	38,407	5.61%		Connection Error	551,179	6.50%
	Malformed Cert	5,275	0.77%		Malformed Cert	194,390	2.29%

Table 5: Detection result comparison between consistency-based heuristics and *CERTainty*—We view the presence of a valid certificate as a strong signal of correct DNS resolution. Invalid certificates that do not match with control, as well as the presence of blockpages are strong signals of DNS manipulation. We use "malformed cert" to denote the invalid certificates that match with invalid control certificate. The consistency-based heuristics include AS number and name, HTTP hash, certificate hash, and PTR lookups.

a general CDN certificate when no domain is specified. Our primary analysis shows that among all the DNS resolution pairs that return a valid certificate when queried with the domain as SNI, 63.89% return general CDN certificates with mismatched hostnames, or no certificates at all when requested without SNI. This indicates that the certificate metrics proposed in previous work introduce FPs into the system. Moreover, Iris reported poor performance (40% to 70% accuracy) in using certificates to detect correct DNS resolution, attributing the performance issue to widespread misconfiguration of TLS servers. Since 2018, more servers have adopted TLS, and hence we find certificate validation to be more useful [61].

Page information: Previous work has incorporated information extracted from the page fetched from resolved IPs, either using the page length [3] or identifying blockpages [20, 45]. The presence of blockpages, like certificates, not only signals the existence of DNS manipulation but also pinpoints the actors. However, among all the DNS manipulation detected by *CERTainty*, 82.39% observe invalid certificates without blockpages. Therefore, blockpage information alone is not enough. Moreover, we discover and publish fingerprints for 226 new blockpages that are not covered by previous open-sourced databases. In later sections (§6), we show how blockpage fingerprints and properly designed certificate validation in tandem can shed light on actors of DNS manipulation.

5.2 Comparing with Consistency-based Heuristics

As shown in Table 1, all state-of-the-art DNS manipulation detection systems incorporate a “test vs. control” strategy—comparing potentially censored responses and their metadata with responses from a set of trusted resolvers. Unfortunately, due to an increasingly complex Internet ecosystem, subsequent requests to a single domain even from unmanipulated networks may return different IPs, or different site content, due to complexities in load balancing, CDN deployments, or geo-targeted content serving. To demonstrate how these errors can occur and how our technique compares, we compare against four popular consistency-based techniques: (1) HTTP and Certificate hash matching, (2) AS matching, (3) PTR matching, and finally, (4) statistical thresholding.

When the IP resolved during the test request matches with at least one IP in the control set, we believe that there is no DNS manipulation in place, which is the case in 70.22% of our DNS resolutions. The aforementioned consistency-based heuristics help cover instances where control IPs fail to account for different points of presence of a given domain. Thus, we find that 28.78% of test requests did not return an IP in our control set. We apply the above consistency-based heuristic comparisons to these 28.78% of cases using metadata from Censys and Maxmind [15, 40].

Overall, we observe that 9.70% of true manipulated responses—having an invalid certificate or matching a blockpage fingerprint—are erroneously tagged as correct resolution using consistency-based comparisons i.e. 9.70% of the cases are false negatives. Moreover, a staggering number of 72.45% DNS resolutions that are tagged as “manipulation“ by consistency-based heuristics are false positives.

Investigating false negatives in consistency-based heuristics: We provide a breakdown of the false negatives of each of the consistency-based heuristics below (see Table 4 for an overview).

AS and PTR (CDN) Matching: Prior work heavily relies on additional consistency checks based on AS details (name or number) and also performs PTR lookups to check whether the IP address served sits in the same CDN or cloud provider as control IPs. Unfortunately, these metrics are frequently flawed, acting as the major source of false negatives. In our experiments, among all the false negative results, 87.02% have a matching control AS name or AS name. Almost all of the false negatives have matching control CDN names, as shown in Table 4. This is because some filtering device vendors, like Securly and Infoblox, serve their blockpages on IPs in big CDNs (e.g. Amazon), which would then be erroneously flagged as *not* DNS manipulation.

HTTP and Certificate Hash Matching: The false negatives introduced by control HTTP hash and certificate hash matching are because previously proposed techniques do not perform sanity checks for the control contents. Conceptually, matching certificate hashes or HTTP hashes (fetched from Censys [15]) between the control and test IP addresses can be strong signals of correct DNS resolution. However, some CDNs return a general CDN certificate and an error page when issued a malformed request or a request for an IP address instead of a domain name.

We observe that our control resolutions sometimes point to these CDNs, and the HTTP and certificate hashes that are stored in these

ASN	AS Owner	Count	Percentage	Type
AS3303	Swisscom	86,115	13.63%	CDN
AS9498	Airtel	82,099	13.00%	CDN
AS20940	Akamai	63,592	10.07%	CDN
AS1299	Arelion	33,763	5.35%	CDN
AS139341	Aceville Pte	18,183	2.88%	Cloud Provider
AS54113	Fastly	16,153	2.56%	CDN
AS24940	Hetzner	12,524	1.98%	Cloud Provider
AS9121	Türk Telekom	11,815	1.87%	Telecom
AS9002	RETN	10,380	1.64%	Telecom

Table 7: Characterizing false positives of consistency-based heuristics through AS distribution—The distribution of the top 10 ASes whose IPs are misclassified as manipulation by consistency-based hueristics.

cases are of these error pages and general CDN certificates. These sometimes match with the HTTP and certificate hashes obtained during the test resolution, even though the resolution itself is incorrect. These cases can arise when the manipulated content is hosted on the same network as the legitimate content. We confirm using *CERTainty* that these resolutions are incorrect based on sending a query for the `(IP, domain)` pairs. Among all the false negatives observed, 3.12% see an HTTP hash match and 3.85% see a certificate hash match.

AS Consistency Thresholding: For IPs that are not in the same AS as control IPs, other work has proposed using fine-tuned thresholds [24, 45] to observe how many websites resolve to the same IP. For example, if a set of websites resolve to a single IP from test vantage points, but resolve to IPs in more than θ ASes from the control nodes, then the test responses for those websites are flagged as DNS manipulation.

To evaluate how effective this thresholding scheme is when compared to checking the certificate validity, we plot the results for DNS resolutions at the global, China, and Indonesia level at each threshold (Figure 4). We find that such a threshold method is only capable of identifying a specific kind of DNS manipulation (e.g. the DNS manipulation in Indonesia), but omits others (e.g. the DNS manipulation in China). Overall, the false negative rate on the global scale is over 50% and increases as the false positive rate drops. Like other consistency heuristics, we find this thresholding metric to be fragile and introduce errors into DNS manipulation results.

Overall, 9.7% of the responses marked as legitimate DNS resolutions by the combination of the above consistency-based metrics return an *invalid* certificate or match a blockpage as detected by *CERTainty*. As shown in Table 4, AS and CDN control matching is the major source of false negatives. Despite the community recognizing AS matching as the most powerful consistency heuristic (see Table 1), it cannot detect filtering devices hosting their blockpages on major CDNs.

***Investigating False Positives in consistency-based heuristics*:** To investigates cases where consistency-based heuristics falsely label correct DNS resolutions as manipulated, we compare our results holistically against all the aforementioned metrics taken in tandem, as this is the approach taken in previous work to detect DNS manipulation [20, 53, 66]. In particular, we investigate how the collective determination of consistency-based heuristics (i.e., manipulated or unmanipulated) differ from *CERTainty*'s determinations.

OpenDNS IP	Hostname
146.112.61.105	hit-botnet.opendns.com
146.112.61.106	hit-adult.opendns.com
146.112.61.107	hit-malware.opendns.com
146.112.61.108	hit-phish.opendns.com
others	hit-block.opendns.com

Table 8: IPs owned by OpenDNS detected by *CERTainty*.

As shown in Table 5, we observe that the 72.45% of the DNS responses tagged as manipulated (i.e. do not match with *any* control data) contain IP addresses that host a valid certificate for the queried domains, which we consider as false positives.

In investigating these false positives, we uncover that the vast majority of IPs that were tagged as DNS manipulation are hosted by CDNs and ISPs (Table 7). However, these CDNs and ISPs are not known adversaries—rather, they may deploy highly distributed resolvers that return IPs based on a number of different decisions (e.g. anycast, load balancing) which the consistency metrics proposed by previous work do not adequately capture.

In addition, there are two other reasons for the presence of these false positives:

(1) The coverage of control resolvers is always limited. Although we report 70.22% effectiveness of using matching control IPs to identify unmanipulated responses, the control groups fail to cover all the points of presence.
(2) All metadata information used for consistency checks (e.g., AS information, HTTP hashes, and certificate validation) are collected from auxiliary databases (Censys and Maxmind) in our work as well as previous work [53, 66] and may be incomplete. For the IPs returned by test resolvers in our measurements, only 30.3% have a certificate hash, 93% have an HTTP hash, and 99% have a matching AS name and AS number.

In summary, by comparing *CERTainty*'s findings across prior work's consistency metrics, we demonstrate how identifying DNS manipulation via IP metadata matching can provide fragile and sometimes incorrect results—72.45% of the manipulated detected by the consistency metrics are false positive, and 9.70% manipulated DNS responses are omitted. In contrast, certificate validation and blockpage matching provide a clear improvement in accuracy.

6 FINDINGS

In this section, we describe the key findings we observe from investigating DNS manipulation with certificate validation. *CERTainty* discovers commercial filtering product deployment in 52 countries, as well as ISP-level DNS manipulation in 26 countries, with a wide diversity of DNS manipulation deployment strategies. 82.39% of the invalid certificates we detect come without a blockpage. To the best of our knowledge, this is the first report of the implementers of DNS manipulation on a global scale. Through the lens of certificate

	Product	Origin	Block Page	Root Cert	Country of Deployment
Observed in one country	Canadian Shield	CA	🔴	▲	CA
	WebTitan	US	🔴	🔺	US
	OneDNS	CN	🔴	▲	CN
	JusprogDNS	DE	🔴	🔺	DE
	Infoblox	US	🔴	🔺	US
	NextDNS	US	🔴	🔺	US
	Comodo	US	🔴	▲	US
	Zyxel	CH	🔴	🔺	CH
	WatchGuard	US		🔺	US
	Securly	US	🔴	🔺	US
Observed in multiple countries	OpenDNS (Cisco)	US	🔴	🔺	AR, AU, BR, CA, CL, CN, CR, CZ, DE, ES, FR, GR, ID, IE, IN, IT, JP, KR, KZ MX, NZ, RO, SE, SK, US, ZA
	AdguardDNS	CA	🔴	▲	GB, BY, CY, FR, ID, LV, NZ, RU
	SafeDNS	US	🟥	🔺	AU, NL, US
	Kewoen	DE		🔺	AU, DE, FR, GB, JP, NL, US
	SkyDNS	RU	🟥	🔺	RU, UA, KZ
	CloudVeil	US	🔴	🔺	CA, US
	Fortinet	US		▲	AR, AT, AU, BD, BF, BR, CA, CH, CL, CN, CZ, DE, DK, FR, GB, HK, ID, IN, IQ, IT, JP, KR, KW, MR, MY, NL, PH, PL, SV, TH, TR, TT, TW, US

Table 9: Location of origin and deployment of different filtering products identified via certificate validation and blockpage matching - (1) For the Blockpage column, a red square indicates a legal blockpage, a red circle indicates a blockpage that does not contain legal information. (2) For the Root Cert column, a green triangle indicates a trusted root CA, a red triangle indicates an untrusted root CA.

validation and blockpage matching, we are not only able to detect signals of DNS manipulation but also pinpoint the actors.

6.1 Identifying Filtering Product Vendors

CERTainty identifies 17 DNS manipulation filtering product vendors deployed in 52 countries, as shown in Table 9. Most (94.11%) commercial filtering devices return an IP hosting (configurable) blockpages. Vendors deploy different strategies for DNS manipulation. For example, certificate chains returned by IPs owned by Cira, OneDNS, AdguardDNS, and Fortinet have a trusted root CA e.g. DigiCert and ZeroSSL. In this case, the common name of the certificates is usually issued for the product website (e.g. `*.onedns.net`). Other products attempt to perform a man-in-the-middle i.e. the leaf certificate has a matching hostname with the queried domain, yet the root CA is not trusted by major browsers.

We discover 10 vendors whose products are only seen in the country of origin. We also observe a concerning pattern where DNS manipulation vendors export their products to countries without such technology, similar to findings from previous work on the spread of HTTP filtering devices [12, 59]. SkyDNS, a vendor based in Russia, claims that their technology is already used in more than 45 countries, advertised as “Solutions for country wide content filtering and network security” [64]. *CERTainty* captures blockpages with 451 status code (“Unavailable For Legal Reasons”) from SkyDNS products and SafeDNS, implying that they are targeting ISPs and governments as potential consumers.

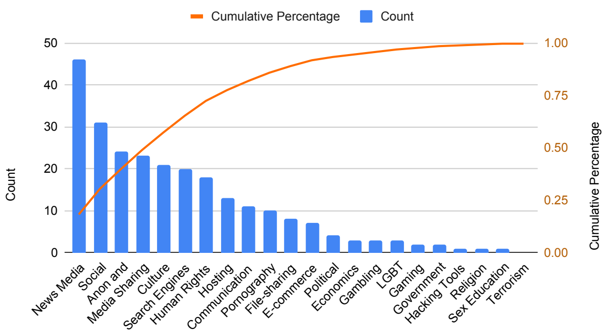


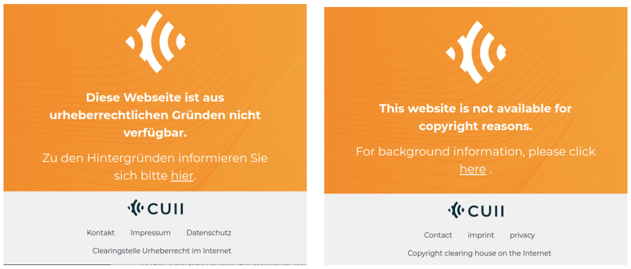
Figure 5: Category Distribution of domains blocked in China - *CERTainty* detects DNS manipulation in China via certificate validation.

We observe that 7 commercial filtering products with multinational coverage return a common set of IP addresses in all countries, indicating that these IP addresses are managed centrally. For example, among all the DNS responses from resolvers in 34 countries that are tampered by Fortinet, only one IP is observed (`208.91.112.55`), which sits in AS40934 (Fortinet Inc). Interestingly, OpenDNS, Cisco’s DNS subsidiary, owns a small pool of anycast IPs whose hostname indicates the exact content they are blocking (Table 8). This observation can potentially be used in DNS manipulation circumvention. If the middleboxes are *injecting* the tampered IPs, the DNS client can be modified to discard a known pool of DNS manipulation product IPs to wait for the correct resolution to arrive.

6.2 Identifying ISP DNS Manipulation

CERTainty detects DNS manipulation on the ISP level in 26 countries via certificate validation, ranging from previously well-studied countries in Internet censorship e.g. Russia [49, 60], to countries that previous research in Internet censorship did not investigate in depth, e.g. Indonesia, Nepal, Thailand, and Romania. We also see ISPs performing DNS manipulation in countries that Freedom House classified as “Free” [30], such as Germany, Greece, and Denmark. ISPs in these countries usually return a blockpage indicating blocking of copyright-infringing domains (as shown in Figure 6). In “Not Free” countries such as China, Russian, and Iran, the blocked categories include news media, circumvention tools, social media, and other major services e.g. Google and Wikipedia (as shown in Figure 5). *CERTainty*’s dataset also covers signals of DNS manipulation implemented by companies, universities, and other organizations.

We observe heterogeneity in ISP-level DNS manipulation even for ISPs within the same country. As shown in Table 10, ISPs can



Diese Webseite ist aus urheberrechtlichen Gründen nicht verfügbar.

Zu den Hintergründen informieren Sie sich bitte hier.

CUII

Kontakt Impressum Datenschutz

Clearingstelle Urheberrecht im Internet

This website is not available for copyright reasons.

For background information, please click here .

CUII

Contact imprint privacy

Copyright clearing house on the Internet

Figure 6: Copyright infringement blockpage from CUII—a joint initiative of affected German industry associations and ISPs [28]

either use services like Let's Encrypt, DigiCert, or Sectigo to issue certificates trusted by major browsers for their blockpages, or simply issue self-signed certificates, which are easy to create and do not involve any financial cost. Most ISPs return a blockpage for blocked contents, either citing the laws and regulations that legitimize such blocking or simply state the access is forbidden or denied in their local languages. However, we observe that some of them just return a blank page or a default setup page of OpenResty. The common name and issuer field in the certificates issued by ISPs are informative of DNS manipulation as well. Some indicate the ISPs are the issuers of the certificates, others even indicate that the certificates are issued by the ISP for the purpose of DNS manipulation.

For example, in Russia, we see tampered IPs in 16 different ASes returned by ISP-owned resolvers. *CERTainty* identifies 31 unique blockpage fingerprints in Russia, revealing the decentralized nature of Russia's national DNS manipulation [60]. We see the presence of badly configured certificates that have no information in all the fields except the effective date, expiration date, and the country of the issuer (“`RU`“). We also observe carefully configured certificates that specify both the ISP and the purpose of the certificate in the common name e.g. `forbidden.citytelecom.ru`, issued by an ISP based in Moscow. DNS manipulation by ISPs can be quite obscure if the implementer chooses to not return a blockpage and does not configure an informative certificate. For example, in Romania, we see a certificate with an IP address in the common name (`213.177.28.90`) returned with the HTTPS page. The IP address in the certificate common name hosts a blockpage stating that the access to the requested website is blocked by the decision of the Supervisory Committee of the O.N.J.N, the gambling regulation institution of Romania. Some ISPs only return a default webpage like “Welcome to nginx¡‘ along with their ISP certificate. Therefore, checking blockpage matching and certificates in tandem helps us to have a more holistic view of overt DNS manipulation.

We also discover ISP-level DNS manipulation via HTTP-only blockpage matching in countries including Russia, Indonesia, Turkey, Poland, Italy, Romania, India, Columbia, Belgium, Philippines, Mexico, Australia, Nepal, and Ukraine. Examples of ISP-level HTTP blockpages without certificates are shown in Fig. 8 in Appendix A.3. In total, *CERTainty* discovers ISP-level DNS manipulation in 26 countries.

Country	AS number of returned IPs	Leaf Cert	Block Page	Root Cert
Russia	AS12616, AS44347, AS44587, AS49505, AS34241	🟥	🟥	▲
	AS25549, AS31483, AS34757	🔴	🔴	▲
	AS12389, AS50466		🟥	🔺
	AS42071, AS42071	🔴		🔺
	AS57571, AS43287, AS49469	🟥	🟥	🔺
	AS8395	🔴	🟥	🔺
Ukraine	AS42546	🔴	🟥	▲
	AS42546	🔴	🔴	▲
Indonesia	AS58396, AS45287, AS45287, AS45287, AS38758	🔴	🔴	🔺
	AS9341, AS9341, AS5578, AS9341		🔴	🔺
	AS16276, AS141626, AS141626, AS7713	🔴	🔴	🔺
	AS58495, AS132634	🟥		▲
	AS140413, AS136873	🟥	🟥	▲
	AS56241	🔴	🔴	▲
Nepal	AS63991	🔴		▲
	AS140973	🟥	🔴	▲
Thailand	AS23969	🟥		▲
Singapore	AS3758, AS3758	🔴		🔺
Belarus	AS6697		🟥	🔺
Lithuania	AS212531	🔴	🟥	▲
Romania	AS31313		🟥	🔺
	AS12302			🔺
Belgium	AS2611	🔴	🔴	▲
	AS5432, AS8717	🔴		🔺
Denmark	AS35158	🔴	🔴	▲
Italy	AS29050	🟥		🔺
Columbia	AS35158	🔴	🟥	▲
Greece	AS6799	🟥		🔺
Switzerland	AS3303	🔴	🟥	▲
Germany	AS24940		🟥	🔺
Australia	AS16509	🟥	🟥	▲

Table 10: Countries where *CERTainty* detects ISP-level DNS manipulation via certificate validation—(1) For the Leaf Cert column, a red square indicates that the ISP is specified as the issuer and the certificate is issued for blocking. A red circle indicates only that the certificate is issued by the ISP. (2) For the Blockpage column, a red square indicates a legal blockpage, and a red circle indicates a blockpage that does not contain legal information. (3) For the Root Cert column, a green triangle indicates a trusted root CA, and a red triangle indicates an untrusted root CA.

6.3 Identify Covert DNS manipulation

From analyzing the heterogeneity of DNS manipulation practice by commercial products and ISP deployment, we learn that it is

Country/Region	Hong Kong	Singa-pore	Bangladesh, South Korea, Indonesia, Myanmar, Thailand
Overlap	85.43%	85.96%	100%
Domains	198	174	< 5

Table 11: Countries/Regions that are potentially affected by China's censorship leakage—The overlap indicates the fraction of `(domain, resolved IP)` **returned by affected resolvers outside China that overlap with Chinese DNS manipulation.**

important to integrate both the information inferred from blockpages and certificates. In the worst case, a adversary can choose to issue a non-informative certificate with no blockpages, making it very hard to determine the implementer and purpose of DNS manipulation, giving the adversary itself plausible deniability of implementing Internet blocking. In this subsection, we will investigate a case of covert DNS manipulation discovered by certificate validation.

Among all the invalid certificates *CERTainty* detected, 82.39% come without a blockpage. About 39.17% of the invalid certificates do not contain information about filtering product vendors or ISPs, making it more covert. The certificates in this category have trusted root CAs and mismatching hostnames, coming with an HTTPS page with a client error status code (400+), indicating that the queried HTTPS servers either can not find the resource the user requested, or think it is a bad request. 98.66% of those IPs are returned by DNS resolvers in China. The returned IPs belong to a huge IP pool located in ASes owned by Facebook, Twitter, Cloudflare, and other blocked CDN providers, which confirms the finding in previous research [25, 26, 39, 46, 74] that China is injecting IP addresses belonging to popular US companies. The categories of domains blocked in China are shown in Fig. 5. Resolvers in 14 other countries and regions share this behavior, seven of which contain at least one resolver that sees a complete overlap of the same DNS manipulation methodology as the Chinese GFW, as shown in Table 11. China's DNS injections are sometimes cached by resolvers outside China, despite the administrators of those DNS resolvers having no intention to implement DNS manipulation [25, 43, 47].

0.6% of the certificates that come with a status code 400+ are issued by Cira, a DNS firewall "to block access to malicious websites" [10], shown in Table 9. As another case of covert DNS manipulation, we observe countries such as Iran utilizing private IPs to block sensitive content. This form of DNS manipulation is more opaque than the previously discussed cases. It is very hard to determine whether the DNS manipulation is intentional when no blockpages are served and the traffic is diverged to the IPs that are not owned by the adversaries. Judging if the blocking is intentional is a hard task in covert DNS manipulation, but by checking the validity of certificates returned, we are able to obtain the users' perspective and understand if the right resource is hosted on the returned IPs.

6.4 Case Study: IPs With No Certificates

Investigating cases where test revolvers fail for both HTTP and HTTPS requests produces indicators of misconfigured resolvers (i.e. all domains tested resolved to the same IP and subsequently fail for both HTTP and HTTPS) but also indicators of revolvers configured for specific domain blocking. For example, we discover 83 Russian resolvers that assign between 20 and 114 domains to the IP `62.33.207.197`. The domains assigned to this IP by the resolver include `bbc.com`, `bridges.torproject.org`, and `psiphon.ca`. Upon investigation, we discover that port 80 and port 443 on this IP address are closed; the only open port is port 444, which returns a Russian blockpage (Figure 7). *CERTainty* does not aim to scan all the potential open ports of the returned IPs. However, by filtering out resolvers that return the same IP for multiple queried domains, we obtain potential signals of DNS manipulation that can be confirmed by further analysis.

7 LIMITATIONS

In this paper, we only consider DNS manipulation on the conventional port 53 (as well as page fetch from port 80 for HTTP and port 443 for HTTPS). However, previous work has shown that DNS manipulation can happen on other ports as well [6]. Moreover, we do not consider the rare possibility of rogue certificate authorities [57, 58, 69]. It is possible for nation-state actors to conduct such silent MitM attacks while evading our detection. We rely on major browsers such as Mozilla Firefox and Google Chrome to remove such root CAs from their trust lists.

Censors could employ unreachable IPs to prevent users from accessing the domains they request, using either (1) private IPs or (2) public IPs that do not host anything on port 80 or port 443. We identify DNS manipulation performed using private IPs, but choose to mark the second case as *unknown*. Future work can manually investigate those IPs to identify whether these cases are DNS manipulation. A case study of such an investigation can be found in Appendix A.2. Moreover, for the case where the IPs only host an HTTP page, we do not attempt to identify real pages by comparing page content. Instead, we match HTTP pages with blockpage fingerprints. Web services often have country-specific dynamic content which could lead to inaccuracies.

While this study includes 6 months of data from Censored Planet, it is worth noting that further longitudinal analysis of DNS data has the potential to capture useful signals regarding emerging patterns and trends in DNS manipulation. Diversifying the geolocation of vantage points for DNS resolution measurement and HTTP(S) page fetch can yield data that more accurately reflect the user experience of DNS censorship. Such an approach can provide a more comprehensive understanding of the global patterns and impact of DNS manipulation. We hope future research will analyze the discrepancies between measurements obtained from different vantage points. Such analysis can enable researchers to gain a more complete picture of global content delivery and to identify instances of server-side blocking.

8 CONCLUSION

Developing heuristics to accurately detect DNS manipulation on a global scale is challenging. State-of-the-art heuristics introduced by previous work that identify shared infrastructure—though intuitive—are error-prone given the advancements in Internet infrastructure such as CDNs. We discover that 72.45% of the manipulated DNS responses identified by the current state-of-the-art are false positives. By taking one step forward to fetch the HTTP(S) pages hosted on IPs returned by resolvers, *CERTainty* simulates the users' perspective to understand the accessibility of requested resources. By leveraging certificate validation and blockpage matching, *CERTainty* identifies 17 TLS proxy vendors deployed in 52 countries, as well as 26 countries with ISP-level DNS manipulation. From *CERTainty*'s dataset, we construct 226 unique blockpage fingerprints for previously unknown blockpages. Our techniques and the curated blockpage fingerprints are open-sourced. We have collaborated with Censored Planet [66], an open censorship measurement platform, to integrate our techniques into its functioning, and we are actively working on integrating our techniques into other measurement platforms such as OONI [20]. We hope our techniques enable accurate detection of DNS manipulation and improve the quality of open-access data provided to the Internet freedom community.

9 ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their helpful comments on the paper. We also thank Will Scott, David Wang, and Daniel Liu for their valuable feedback on DNS measurements. This work was supported by the Defense Advanced Research Projects Agency under Agreement No. HR00112190127 and a Bureau of Democracy, Human Rights and Labor (DRL) Grant (No. SLMAQM20GR2132).

REFERENCES

[1] Mustafa Emre Acer, Emily Stark, Adrienne Porter Felt, Sascha Fahl, Radhika Bhargava, Bhanu Dev, Matt Braithwaite, Ryan Sleevi, and Parisa Tabriz. Where the wild warnings are: Root causes of chrome https certificate errors. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1407–1420, 2017.

[2] Giuseppe Aceto, Alessio Botta, Antonio Pescapé, M Faheem Awan, Tahir Ahmad, and Saad Qaisar. Analyzing internet censorship in pakistan. In *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)*, pages 1–6. IEEE, 2016.

[3] Giuseppe Aceto, Alessio Botta, Antonio Pescapè, Nick Feamster, M Faheem Awan, Tahir Ahmad, and Saad Qaisar. Monitoring internet censorship with ubica. In *International Workshop on Traffic Monitoring and Analysis*, pages 143–157. Springer, 2015.

[4] Devdatta Akhawe, Johanna Amann, Matthias Vallentin, and Robin Sommer. Here's my cert, so trust me, maybe? understanding tls errors on the web. In *Proceedings of the 22nd international conference on World Wide Web*, pages 59–70, 2013.

[5] Simurgh Aryan, Homa Aryan, and J Alex Halderman. Internet censorship in iran: A first look. In *3rd USENIX Workshop on Free and Open Communications on the Internet (FOCI 13)*, 2013.

[6] Abhishek Bhaskar and Paul Pearce. Many roads lead to rome: How packet headers influence DNS censorship measurement. In *31st USENIX Security Symposium (USENIX Security 22)*. USENIX Association, August 2022.

[7] Kevin Bock, Yair Fax, Kyle Reese, Jasraj Singh, and Dave Levin. Detecting and evading {Censorship-in-Depth}: A case study of {Iran's} protocol whitelister. In *10th USENIX Workshop on Free and Open Communications on the Internet (FOCI 20)*, 2020.

[8] T. Brisco. Rfc1794: Dns support for load balancing, 1995. https://datatracker.ietf.org/doc/html/rfc1794.

[9] Abdelberi Chaabane, Terence Chen, Mathieu Cunche, Emiliano De Cristofaro, Arik Friedman, and Mohamed Ali Kaafar. Censorship in the wild: Analyzing internet filtering in syria. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, pages 285–298, 2014.

[10] Cira. Cira canadian shield, 2022. https://www.cira.ca/cybersecurity-services/canadian-shield.

[11] Molly Contreras. Cisco umbrella | leader in cloud cybersecurity & sase solutions, 2022. https://umbrella.cisco.com/.

[12] Jakub Dalek, Bennett Haselton, Helmi Noman, Adam Senft, Masashi Crete-Nishihata, Phillipa Gill, and Ronald J Deibert. A method for identifying and confirming the use of url filtering products for censorship. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 23–30, 2013.

[13] Ronald J Deibert and Masashi Crete-Nishihata. Global governance and the spread of cyberspace controls. *Global Governance*, 18:339, 2012.

[14] Trinh Viet Doan, Irina Tsareva, and Vaibhav Bajpai. Measuring dns over tls from the edge: adoption, reliability, and response times. In *International Conference on Passive and Active Network Measurement*, pages 192–209. Springer, 2021.

[15] Zakir Durumeric, David Adrian, Ariana Mirian, Michael Bailey, and J Alex Halderman. A search engine backed by internet-wide scanning. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pages 542–553, 2015.

[16] Zakir Durumeric, Eric Wustrow, and J Alex Halderman. {ZMap}: Fast internet-wide scanning and its security applications. In *22nd USENIX Security Symposium (USENIX Security 13)*, pages 605–620, 2013.

[17] Roya Ensafi, Philipp Winter, Abdullah Mueen, and Jedidiah R Crandall. Analyzing the great firewall of china over space and time. *Proc. Priv. Enhancing Technol.*, 2015(1):61–76, 2015.

[18] Victor Le Pochat et al.. Tranco: A research-oriented top sites ranking hardened against manipulation, 2022. https://tranco-list.eu/.

[19] Leonid Evdokimov and Vasilis Ververis. Identifying cases of DNS misconfiguration: Not quite censorship, 2017. https://ooni.org/post/not-quite-network-censorship/.

[20] Arturo Filasto and Jacob Appelbaum. OONI: Open Observatory of Network Interference. In *USENIX Workshop on Free and Open Communications on the Internet (FOCI)*, 2012.

[21] King-wa Fu, Chung-hong Chan, and Michael Chau. Assessing censorship on microblogs in china: Discriminatory keyword analysis and the real-name registration policy. *IEEE internet computing*, 17(3):42–50, 2013.

[22] Gamban. Block access to gambling websites and apps worldwide, 2022. https://gamban.com/.

[23] Sebastián García, Karel Hynek, Dmtrii Vekshin, Tomáš Čejka, and Armin Wasicek. Large scale measurement on the adoption of encrypted dns. *arXiv preprint arXiv:2107.04436*, 2021.

[24] Phillipa Gill, Masashi Crete-Nishihata, Jakub Dalek, Sharon Goldberg, Adam Senft, and Greg Wiseman. Characterizing web censorship worldwide: Another look at the opennet initiative data. *ACM Transactions on the Web (TWEB)*, 9(1):1–29, 2015.

[25] Nguyen Phong Hoang, Sadie Doreen, and Michalis Polychronakis. Measuring {I2P} censorship at a global scale. In *9th USENIX Workshop on Free and Open Communications on the Internet (FOCI 19)*, 2019.

[26] Nguyen Phong Hoang, Arian Akhavan Niaki, Jakub Dalek, Jeffrey Knockel, Pellaeon Lin, Bill Marczak, Masashi Crete-Nishihata, Phillipa Gill, and Michalis Polychronakis. How great is the great firewall? measuring china's {DNS} censorship. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 3381–3398, 2021.

[27] J. Hodges. Rfc6125: Representation and verification of domain-based application service identity within internet public key infrastructure using x.509 (pkix) certificates in the context of transport layer security (tls) section-6.4, 2011. https://datatracker.ietf.org/doc/html/rfc6125.html.

[28] Jochen Homann. Online copyright clearance system arranges block of streaming site, 2021. https://tinyurl.com/22w8vdmv.

[29] Amir Houmansadr, Giang TK Nguyen, Matthew Caesar, and Nikita Borisov. Cirripede: Circumvention infrastructure using router redirection with plausible deniability. In *Proceedings of the 18th ACM conference on Computer and communications security*, pages 187–200, 2011.

[30] Freedom House. Freedom house: Global freedom status, 2022. https://freedomhouse.org/explore-the-map.

[31] Ben Jones, Tzu-Wen Lee, Nick Feamster, and Phillipa Gill. Automated detection and fingerprinting of censorship block pages. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, pages 299–304, 2014.

[32] Keweon. Keweon, secure and reliable open internet, 2022. https://websrv.keweon.center/.

[33] Christian Kreibich, Nicholas Weaver, Boris Nechaev, and Vern Paxson. Netalyzr: Illuminating the edge network. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, pages 246–259, 2010.

[34] Deepak Kumar, Zhengping Wang, Matthew Hyder, Joseph Dickinson, Gabrielle Beck, David Adrian, Joshua Mason, Zakir Durumeric, J Alex Halderman, and Michael Bailey. Tracking certificate misissuance in the wild. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 785–798. IEEE, 2018.

[35] Citizen Lab and Others. Url testing lists intended for discovering website censorship, 2014. https://github.com/citizenlab/test-lists.

[36] Jyh-An Lee and Ching-Yi Liu. Real-name registration rules and the fading digital anonymity in china. *Wash. Int'l LJ*, 25:1, 2016.

[37] Graham Lowe, Patrick Winters, and Michael L Marcus. The great dns wall of china. *MS, New York University*, 21(1), 2007.

[38] Chaoyi Lu, Baojun Liu, Zhou Li, Shuang Hao, Haixin Duan, Mingming Zhang, Chunying Leng, Ying Liu, Zaifeng Zhang, and Jianping Wu. An end-to-end, large-scale measurement of dns-over-encryption: How far have we come? In *Proceedings of the Internet Measurement Conference*, pages 22–35, 2019.

[39] Bill Marczak, Nicholas Weaver, Jakub Dalek, Roya Ensafi, David Fifield, Sarah McKune, Arn Rey, John Scott-Railton, Ron Deibert, and Vern Paxson. An analysis

of {China's}{"Great}{Cannon"}. In *5th USENIX Workshop on Free and Open Communications on the Internet (FOCI 15)*, 2015.

[40] Maxmind. Ip geolocation and online fraud prevention, 2022. https://maxmind.com/.

[41] Mimecast. Mimecast dns security, 2022. https://www.mimecast.com/content/dns-security/.

[42] Zubair Nabi. The anatomy of web censorship in pakistan. In *3rd USENIX Workshop on Free and Open Communications on the Internet (FOCI 13)*, 2013.

[43] Hovership Nebuchadnezzar. The collateral damage of internet censorship by dns injection. *ACM SIGCOMM CCR*, 42(3):10–1145, 2012.

[44] NextDNS. Nextdns, the new firewall for the modern internet., 2022. https://nextdns.io/.

[45] Arian Akhavan Niaki, Shinyoung Cho, Zachary Weinberg, Nguyen Phong Hoang, Abbas Razaghpanah, Nicolas Christin, and Phillipa Gill. Iclab: A global, longitudinal internet censorship measurement platform. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 135–151. IEEE, 2020.

[46] Arian Akhavan Niaki, Nguyen Phong Hoang, Phillipa Gill, Amir Houmansadr, et al. Triplet censors: Demystifying great firewall's dns censorship behavior. In *10th USENIX Workshop on Free and Open Communications on the Internet (FOCI 20)*, 2020.

[47] Yevheniya Nosyk, Qasim Lone, Yury Zhauniarovich, Carlos H. Ga n´an, Emile Aben, Samaneh Moura, Giovane C. M. Tajalizadehkhoob, Andrzej Duda, and Maciej Korczy´nski. Intercept and Inject: DNS Response Manipulation in the Wild. In *Proceedings of 2023 Passive and Active Measurement Conference*, Virtual Conference, 2023. Springer.

[48] Federal Chief Information Officers. Does the us government operate a publicly trusted certificate authority?, 2022. https://https.cio.gov/certificates/.

[49] Katherine Ognyanova. In putin's russia, information has you: Media control and internet censorship in the russian federation. In *Censorship, surveillance, and privacy: Concepts, methodologies, tools, and applications*, pages 1769–1786. IGI Global, 2019.

[50] OneDNS. Onedns, 2022. https://onedns.net/.

[51] OpenSSL. Openssl, 2022. https://github.com/openssl/openssl.

[52] Paul Pearce, Roya Ensafi, Frank Li, Nick Feamster, and Vern Paxson. Augur: Internet-wide detection of connectivity disruptions. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 427–443. IEEE, 2017.

[53] Paul Pearce, Ben Jones, Frank Li, Roya Ensafi, Nick Feamster, Nick Weaver, and Vern Paxson. Global measurement of {DNS} manipulation. In *26th USENIX Security Symposium (USENIX Security 17)*, pages 307–323, 2017.

[54] Censored Planet. Blockpage fingerprints, assets-censoredplanet, 2022. https://assets.censoredplanet.org/.

[55] Censored Planet. Dns data - satellite, 2022. https://docs.censoredplanet.org/dns.html.

[56] Ingmar Poese, Steve Uhlig, Mohamed Ali Kaafar, Benoit Donnet, and Bamba Gueye. Ip geolocation databases: Unreliable? *ACM SIGCOMM Computer Communication Review*, 41(2):53–56, 2011.

[57] J Ronald Prins and Business Unit Cybercrime. Diginotar certificate authority breach "operation black tulip". *Fox-IT, November*, page 18, 2011.

[58] Ram Sundara Raman, Leonid Evdokimov, Eric Wurstrow, J Alex Halderman, and Roya Ensafi. Investigating large scale https interception in kazakhstan. In *Proceedings of the ACM Internet Measurement Conference*, pages 125–132, 2020.

[59] Ram Sundara Raman, Adrian Stoll, Jakub Dalek, Reethika Ramesh, Will Scott, and Roya Ensafi. Measuring the deployment of network censorship filters at global scale. In *NDSS*, 2020.

[60] Reethika Ramesh, Ram Sundara Raman, Matthew Bernhard, Victor Ongkowijaya, Leonid Evdokimov, Anne Edmundson, Steven Sprecher, Muhammad Ikram, and Roya Ensafi. Decentralized control: A case study of russia. In *Network and Distributed Systems Security (NDSS) Symposium 2020*, 2020.

[61] Google Transparency Report. Https encryption on the web, 2022. https://transparencyreport.google.com/https/overview?hl=en.

[62] Will Scott, Thomas Anderson, Tadayoshi Kohno, and Arvind Krishnamurthy. Satellite: Joint analysis of {CDNs} and {Network-Level} interference. In *2016 USENIX Annual Technical Conference (USENIX ATC 16)*, pages 195–208, 2016.

[63] Nehad Selaiha. The fire and the frying pan: Censorship and performance in egypt. *TDR*, pages 20–47, 2013.

[64] SkyDNS. Solutions for country wide content filtering and network security, 2022. https://www.skydns.ru/en/.

[65] Ram Sundara Raman, Mona Wang, Jakub Dalek, Jonathan Mayer, and Roya Ensafi. Network measurement methods for locating and examining censorship devices. *The 18th International Conference on emerging Networking EXperiments and Technologies*, 2022.

[66] Ramakrishnan Sundara Raman, Prerana Shenoy, Katharina Kohls, and Roya Ensafi. Censored Planet: An Internet-wide, Longitudinal Censorship Observatory. In *ACM Conference on Computer and Communications Security (CCS)*, 2020.

[67] TitanHQ. Titanhq, cybersecuirty platform delivering a layered security solution., 2022. https://websrv.keweon.center/.

[68] Martino Trevisan, Idilio Drago, Marco Mellia, and Maurizio M Munafo. Automatic detection of dns manipulations. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 4010–4015. IEEE, 2017.

[69] L Tung. Mozilla to china's wosign: We'll kill firefox trust in you after mis-issued github certs. *ZDNet, Sep*, 27, 2016.

[70] Martin Ukrop, Lydia Kraus, Vashek Matyas, and Heider Ahmad Mutleq Wahsheh. Will you trust this tls certificate? perceptions of people working in it. In *Proceedings of the 35th annual computer security applications conference*, pages 718–731, 2019.

[71] Benjamin VanderSloot, Allison McDonald, Will Scott, J Alex Halderman, and Roya Ensafi. Quack: Scalable remote measurement of {Application-Layer} censorship. In *27th USENIX Security Symposium (USENIX Security 18)*, pages 187–202, 2018.

[72] John-Paul Verkamp and Minaxi Gupta. Inferring mechanics of web censorship around the world. In *FOCI*, 2012.

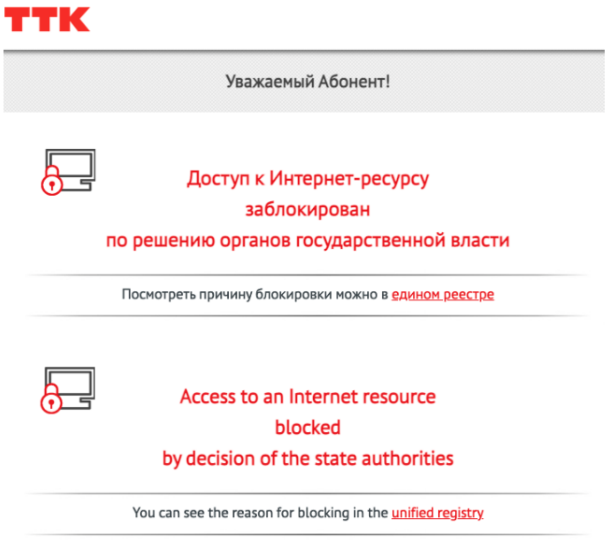
[73] Ben Wagner. Push-button-autocracy in tunisia: Analysing the role of internet infrastructure, institutions and international markets in creating a tunisian censorship regime. *Telecommunications Policy*, 36(6):484–492, 2012.

[74] Zhongjie Wang, Yue Cao, Zhiyun Qian, Chengyu Song, and Srikanth V Krishnamurthy. Your state is not mine: A closer look at evading stateful internet censorship. In *Proceedings of the 2017 Internet Measurement Conference*, pages 114–127, 2017.

[75] Xueyang Xu, Z Morley Mao, and J Alex Halderman. Internet censorship in china: Where does the filtering occur? In *International Conference on Passive and Active Network Measurement*, pages 133–142. Springer, 2011.

[76] Tarun Kumar Yadav, Akshat Sinha, Devashish Gosain, Piyush Kumar Sharma, and Sambuddho Chakravarty. Where the light gets in: Analyzing web censorship mechanisms in india. In *Proceedings of the Internet Measurement Conference 2018*, pages 252–264, 2018.

[77] Kejing Yang. *The Door Is Closed, but Not Locked: China's VPN Policy*. PhD thesis, Georgetown University, 2017.



TTK

Уважаемый Абонент!

Доступ к Интернет-ресурсу
заблокирован
по решению органов государственной власти

Посмотреть причину блокировки можно в едином реестре

Access to an Internet resource
blocked
by decision of the state authorities

You can see the reason for blocking in the unified registry

Figure 7: Russian ISP blockpage hosted on IP `62.33.207.197`, port 444.

A APPENDIX

A.1 Case Study: Nonzero RCODE

By properly filtering out erroneous resolvers, we observe both commercial filtering products and ISPs deploying DNS manipulation by using `RCODE:3 (NXDOMAIN)`. For example, 4 resolvers (0.*.dns.gamban.com) all return `RCODE:3` for exactly 47 domains. The TLD of these resolvers, Gamban, is a commercial filtering product that offers DNS manipulation as a service, which evidently is implemented by returning DNS `NXDOMAIN` [22]. All but 4 of Gamban's 47 blocked sites are gambling domains. The 4 outliers were all circumvention tools: `tunnelbear.com`, `www.ipredator.se`, `torproject.org`, and `bridges.torproject.org`. Although we see the locations of the resolvers to be Great Britain, United States, and Singapore, those appear to be the locations of the Gamban servers, which could be requested by users globally. Previous work [53] found that `NXDOMAIN` is relatively infrequently used for DNS manipulation. However, the existence of such middlebox vendors demonstrates the diversity and evolution of DNS manipulation deployment.

We also see several examples of ISP DNS manipulation utilizing `RCODE:3, NXDOMAIN`. In one example, we see 65 different ROSTELECOM resolvers all of which return `NXDOMAIN` for exactly two domains: `www.facebook.com` and `staticxx.facebook.com`. In Brazil, six resolvers from Telefonica with the TLD `gvt.net.br` return `RCODE:3` for 7 sites:`piratebay.org`, `womenonwaves.org`, and 5 adult content domains. In an example of organizational blocking, we observe 3 resolvers owned by Thai Cyber University using `RCODE:3` to block 6 sites, including `americannaziparty.com` and `nostraightnews.com`. By investigating resolvers that return `NXDOMAIN`, meaningful signals of DNS manipulation emerge from our dataset. We discover the use of `NXDOMAIN` to deploy DNS manipulation, by a diverse set of actors: commercial vendors, ISPs, and organizations like universities and banks.

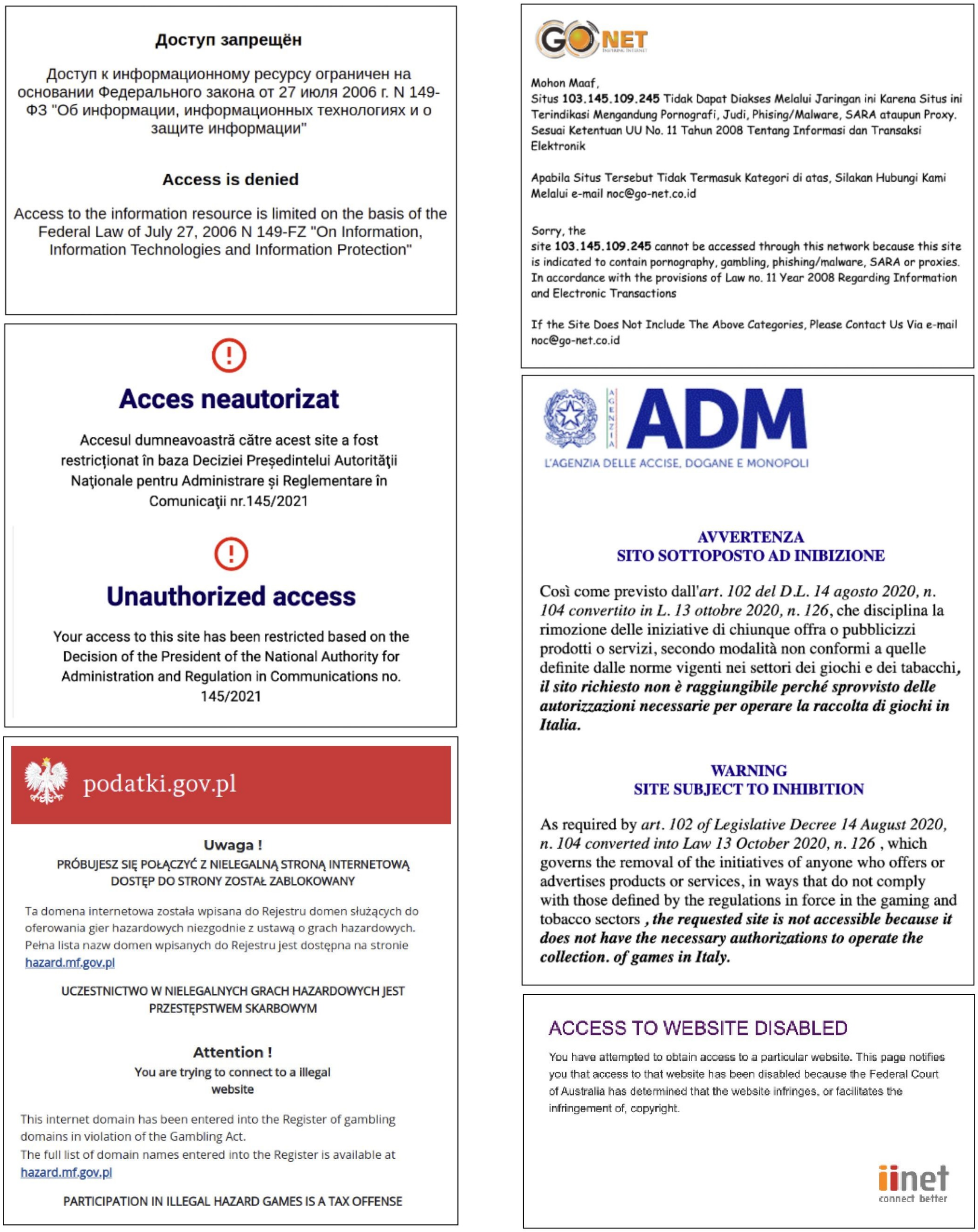
A.2 Annotating DNS responses

More often than not, DNS resolvers return more than one IP for the queried domain. In these cases, a client will need to decide which of the returned IPs to connect to first. The behavior of the DNS client depends on its implementation. Generally, it tries the IPs in the order they were returned by the DNS server in a round robin manner [8].

In most cases of DNS manipulation, we observe that no legitimate IP for the queried domain is included in the DNS responses. In a few very rare cases ($4e^{-5}\%$), we see mixed signals, where IPs hosting a blockpage as well as IPs hosting legitimate content are returned in the same response. This is potentially a case of the collateral damage of DNS poisoning: manipulated DNS records are cached by open resolvers with no intention to block. We mark these cases as unmanipulated in our study.

A.3 HTTP-Only ISP Blockpages

We discover multiple countries with ISP-level DNS manipulation where only HTTP blockpages are returned. Below are 6 examples of such blockpages with corresponding English translation.



Доступ запрещён

Доступ к информационному ресурсу ограничен на основании Федерального закона от 27 июля 2006 г. N 149-ФЗ "Об информации, информационных технологиях и о защите информации"

Access is denied

Access to the information resource is limited on the basis of the Federal Law of July 27, 2006 N 149-FZ "On Information, Information Technologies and Information Protection"

GO NET

Mohon Maaf,
Situs **103.145.109.245** Tidak Dapat Diakses Melalui Jaringan ini Karena Situs ini Terindikasi Mengandung Pornografi, Judi, Phising/Malware, SARA ataupun Proxy. Sesuai Ketentuan UU No. 11 Tahun 2008 Tentang Informasi dan Transaksi Elektronik

Apabila Situs Tersebut Tidak Termasuk Kategori di atas, Silakan Hubungi Kami Melalui e-mail noc@go-net.co.id

Sorry, the
site **103.145.109.245** cannot be accessed through this network because this site is indicated to contain pornography, gambling, phishing/malware, SARA or proxies. In accordance with the provisions of Law no. 11 Year 2008 Regarding Information and Electronic Transactions

If the Site Does Not Include The Above Categories, Please Contact Us Via e-mail noc@go-net.co.id

Acces neautorizat

Accesul dumneavoastră către acest site a fost restricționat în baza Deciziei Președintelui Autorității Naționale pentru Administrare și Reglementare în Comunicații nr.145/2021

Unauthorized access

Your access to this site has been restricted based on the Decision of the President of the National Authority for Administration and Regulation in Communications no. 145/2021

ADM
L'AGENZIA DELLE ACCISE, DOGANE E MONOPOLI

AVVERTENZA
SITO SOTTOPOSTO AD INIBIZIONE

Così come previsto dall'*art. 102 del D.L. 14 agosto 2020, n. 104 convertito in L. 13 ottobre 2020, n. 126*, che disciplina la rimozione delle iniziative di chiunque offra o pubblicizzi prodotti o servizi, secondo modalità non conformi a quelle definite dalle norme vigenti nei settori dei giochi e dei tabacchi, ***il sito richiesto non è raggiungibile perché sprovvisto delle autorizzazioni necessarie per operare la raccolta di giochi in Italia.***

WARNING
SITE SUBJECT TO INHIBITION

As required by *art. 102 of Legislative Decree 14 August 2020, n. 104 converted into Law 13 October 2020, n. 126* , which governs the removal of the initiatives of anyone who offers or advertises products or services, in ways that do not comply with those defined by the regulations in force in the gaming and tobacco sectors ***, the requested site is not accessible because it does not have the necessary authorizations to operate the collection. of games in Italy.***

podatki.gov.pl

Uwaga !
PRÓBUJESZ SIĘ POŁĄCZYĆ Z NIELEGALNĄ STRONĄ INTERNETOWĄ
DOSTĘP DO STRONY ZOSTAŁ ZABLOKOWANY

Ta domena internetowa została wpisana do Rejestru domen służących do oferowania gier hazardowych niezgodnie z ustawą o grach hazardowych. Pełna lista nazw domen wpisanych do Rejestru jest dostępna na stronie hazard.mf.gov.pl

UCZESTNICTWO W NIELEGALNYCH GRACH HAZARDOWYCH JEST PRZESTĘPSTWEM SKARBOWYM

Attention !
You are trying to connect to a illegal website

This internet domain has been entered into the Register of gambling domains in violation of the Gambling Act.
The full list of domain names entered into the Register is available at hazard.mf.gov.pl

PARTICIPATION IN ILLEGAL HAZARD GAMES IS A TAX OFFENSE

ACCESS TO WEBSITE DISABLED

You have attempted to obtain access to a particular website. This page notifies you that access to that website has been disabled because the Federal Court of Australia has determined that the website infringes, or facilitates the infringement of, copyright.

iinet
connect better

Figure 8: ISP HTTP blockpages detected by *CERTainty* - Government blockpages of Russia, Indonesia, Romania, Italy, Poland, and Australia